HamleDT 2.0 License Terms

HamleDT 2.0 (referred to as "HamleDT" in the rest of this document) is a collection of linguistic textual data in multiple languages. It is based on pre-existing data sets ("original treebanks"). Each of the original treebanks has its own license terms and you (the "User") are responsible for complying with the license terms applicable to those parts of HamleDT, which you use. If you do not agree with the license terms, you must stop using HamleDT and destroy all copies of HamleDT data that you have obtained.

You are specifically reminded that some of the original treebanks permit only non-commercial usage.

The additional tree transformations and the software performing the transformations is copyright [Charles University in Prague](), [Faculty of Mathematics and Physics](), [Institute of Formal and Applied Linguistics]() ("we" or the "Provider"). It is provided as-is (without any warranty) and may be freely used, modified and redistributed under the [GNU General Public License version 3.0]() or the [Perl Artistic License version 1.0]().

The treebanks in HamleDT are organized in two groups: "Free" and "Patch". The Free group consists of treebanks whose license terms permit us redistributing them in full. For treebanks in the Patch group, we provide only our transformed annotation, without the underlying texts, lemmas and part of speech tags. If you legally obtained an original treebank, you can use the corresponding patch to transform the treebank to its HamleDT form.

Overview of the "Free" treebanks and their license terms

Lang.	Treebank	Web	License
ar	Prague Arabic Dependency Treebank	http://ufal.mff.cuni.cz/padt	CC BY-NC-SA 3.0
cs	Prague Dependency Treebank	http://ufal.mff.cuni.cz/pdt3.0	CC BY-NC-SA 3.0
da	CoNLL 2006	http://www.buch-kromann.dk/matthias/treebank/	GPL 3
et	Eesti keele puudepank	http://vvv.cs.ut.ee/~kaili/Korpus/puud/	free download
fa	Persian Dependency Treebank	http://dadegan.ir/en/perdt	GPL 3*
fi	Turku Dependency Treebank	http://bionlp.utu.fi/fintreebank.html	CC BY-SA 3.0
grc	Ancient Greek Dependency Treebank	http://nlp.perseus.tufts.edu/syntax/treebank/	CC BY-NC-SA 2.5
la	Latin Dependency Treebank	http://nlp.perseus.tufts.edu/syntax/treebank/	CC BY-NC-SA 2.5
nl	CoNLL 2006 (Alpino)	http://odur.let.rug.nl/~vannoord/trees/	GPL
pt	CoNLL 2006 (Floresta Sintá(c)tica)	http://www.linguateca.pt/Floresta/principal.html	CC BY-NC-SA 3.0
ro	Resurse pentru Gramaticile de Dependenta	http://www.phobos.ro/roric/texts/index ro.html	free download
sv	CoNLL 2006 (Talbanken05)	http://stp.lingfil.uu.se/~nivre/research/Talbanken05.html	research*, cite
ta	Tamil Dependency Treebank	http://ufal.mff.cuni.cz/~ramasamy/tami	CC BY-NC-SA 3.0

Lang.	Treebank	Web	License
		ltb/0.1/	

Overview of the "Patch" treebanks

Lang.	Treebank	Web
bg	BulTreeBank / CoNLL 2006	http://www.bultreebank.org/indexBTB.html
bn	Hyderabad Dependency Treebank / ICON 2010	
ca	AnCora-CA	http://clic.ub.edu/corpus/
de	TIGER Corpus / CoNLL 2009	http://www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/tiger.html
el	Greek Dependency Treebank / CoNLL 2007	
en	Penn Treebank / CoNLL 2007	http://www.cis.upenn.edu/~treebank/
es	AnCora-ES	http://clic.ub.edu/corpus/
eu	Basque Dependency Treebank	
hi	Hyderabad Dependency Treebank / COLING 2012	
hu	Szeged Treebank	http://www.inf.u-szeged.hu/projectdirs/hlt/hu/Treebank/treebank2.html
it	Italian Syntactic-Semantic Treebank / CoNLL 2007	http://www.ilc.cnr.it/viewpage.php/sez=ricerca/id=874/vers=ing
ja	Tübingen Treebank of Spoken Japanese (Tüba-J/S)	http://www.sfs.uni-tuebingen.de/en/ascl/resources/corpora/tueba-js.html
ru	SynTagRus	http://www.ruscorpora.ru/en/search-syntax.html
sk	Slovak Treebank	http://korpus.sk/
sl	Slovene Dependency Treebank	http://nl.ijs.si/sdt/
te	Hyderabad Dependency Treebank / ICON 2010	
tr	METU-Sabanci (ODTÜ-Sabancı) Treebank	http://ii.metu.edu.tr/corpus

Licenses

CC BY-NC-SA 3.0   http://creativecommons.org/licenses/by-nc-sa/3.0/

CC BY-SA 3.0   http://creativecommons.org/licenses/by-sa/3.0/

CC BY-NC-SA 2.5   http://creativecommons.org/licenses/by-nc-sa/2.5/

GPL   http://www.gnu.org/licenses/gpl.html

fa: Persian Dependency Treebank: The download page contained the statement "I will use the

treebank for research purposes only." The "Readme and Licence.txt" file says "only non-commercially". However, the included license is the standard GPL 3 (without any restrictions).

sv: Talbanken05: The download page was saying at the time of retrieval: "The treebank comes with no guarantee but is freely available for research and educational purposes as long as proper credit is given for the work done to produce the material (both in Lund and in Växjö)."

References

The following research papers should be cited to give proper credit to the creators of the original treebanks and to us, the creators of HamleDT.

HamleDT 2.0: Rudolf Rosa, Jan Mašek, David Mareček, Martin Popel, Daniel Zeman, Zdeněk Žabokrtský (2014): HamleDT 2.0: Thirty Dependency Treebanks Stanfordized. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, pp. 2334–2341. ELDA, Reykjavík, Iceland.

ar: Otakar Smrž, Viktor Bielický, Iveta Kouřilová, Jakub Kráčmar, Jan Hajič, Petr Zemánek (2008): Prague Arabic Dependency Treebank: A Word on the Million Words. In *Proceedings of the Workshop on Arabic and Local Languages (LREC 2008)*, pp. 16–23. ELDA, Marrakech, Morocco.

bg: Kiril Simov, Petya Osenova (2005): Extending the Annotation of BulTreeBank: Phase 2. In *The Fourth Workshop on Treebanks and Linguistic Theories (TLT 2005)*, pp. 173–184. Barcelona, Spain.

ca, es: Mariona Taulé, Maria Antònia Martí, Marta Recasens (2008): AnCora: Multilevel Annotated Corpora for Catalan and Spanish. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*. ELDA, Marrakech, Morocco.

cs: Eduard Bejček, Eva Hajičová, Jan Hajič, Pavlína Jínová, Václava Kettnerová, Veronika Kolářová, Marie Mikulová, Jiří Mírovský, Anna Nedoluzhko, Jarmila Panevová, Lucie Poláková, Magda Ševčíková, Jan Štěpánek, Šárka Zikánová (2013): *Prague Dependency Treebank 3.0*. http://hdl.handle.net/11858/00-097C-0000-0023-1AAF-3. Charles University in Prague, ÚFAL, Praha, Czechia.

da: Matthias T. Kromann, Line Mikkelsen, Stine Kern Lynge (2004): *Danish Dependency Treebank*. http://code.google.com/p/copenhagen-dependency-treebank/. København, Denmark.

de: Sabine Brants, Stefanie Dipper, Silvia Hansen, Wolfgang Lezius, George Smith (2002): The TIGER Treebank. In *Proceedings of the Workshop on Treebanks and Linguistic Theories*. Sozopol, Bulgaria.

el: Prokopis Prokopidis, Elina Desipri, Maria Koutsombogera, Harris Papageorgiou, Stelios Piperidis (2005): Theoretical and Practical Issues in the Construction of a Greek Dependency Treebank. In *Proceedings of the Fourth Workshop on Treebanks and Linguistic Theories (TLT 2005)*, pp. 149–160. Barcelona, Spain.

en: Mitchell P. Marcus, Beatrice Santorini, Mary Ann Marcinkiewicz (1993): Building a Large Annotated Corpus of English: The Penn Treebank. In *Computational Linguistics* 19:2, pp. 313–330.

et: Eckhard Bick, Heli Uibo, Kaili Müürisep (2004): Arborest – a VISL-Style Treebank Derived from an Estonian Constraint Grammar Corpus. In *Proceedings of Treebanks and Linguistic Theories (TLT 2004)*.

eu: Itzair Aduriz, María Jesús Aranzabe, Jose Mari Arriola, Aitziber Atutxa, Arantza Díaz de Ilarraza, Aitzpea Garmendia, Maite Oronoz (2003): Construction of a Basque Dependency Treebank. In *Proceedings of the Second Workshop on Treebanks and Linguistic Theories (TLT 2003)*.

fa: Mohammad Sadegh Rasooli, Amirsaeid Moloodi, Manouchehr Kouhestani, Behrouz Minaei-Bidgoli (2011): A Syntactic Valency Lexicon for Persian Verbs: The First Steps towards Persian Dependency Treebank. In *5th Language and Technology Conference (LTC): Human Language Technologies as a Challenge for Computer Science and Linguistics*, pp. 227–231. Poznań, Poland.

fi: Katri Haverinen, Timo Viljanen, Veronika Laippala, Samuel Kohonen, Filip Ginter, Tapio Salakoski (2010): Treebanking Finnish. In *Proceedings of the Ninth International Workshop on Treebanks and Linguistic Theories (TLT9)*, pp. 79–90.

grc, la: David Bamman, Gregory Crane (2011): The Ancient Greek and Latin Dependency Treebanks. In *Language Technology for Cultural Heritage*, pp. 79–98. ISBN 978-3-642-20227-8. Springer, Berlin / Heidelberg, Germany.

it: Simonetta Montemagni, Francesco Barsotti, Marco Battista, Nicoletta Calzolari, Ornella Corazzari, Alessandro Lenci, Antonio Zampolli, Francesca Fanciulli, Maria Massetani, Remo Raffaelli, Roberto Basili, Maria Teresa Pazienza, Dario Saracino, Fabio Zanzotto, Nadia Mana, Fabio Pianesi, Rodolfo Delmonte (2003): Building the Italian Syntactic-Semantic Treebank. In Anne Abeillé (ed.): *Building and Using Parsed Corpora*, pp. 189–210. Kluwer, Dordrecht, Netherlands.

hi, bn, te: Samar Husain, Prashanth Mannem, Bharat Ambati, Phani Gadde (2010): The ICON-2010 Tools Contest on Indian Language Dependency Parsing. In *Proceedings of ICON-2010 Tools Contest on Indian Language Dependency Parsing*. Kharagpur, India.

hu: Dóra Csendes, János Csirik, Tibor Gyimóthy, András Kocsor (2005): The Szeged Treebank. In *Text, Speech and Dialogue (TSD)*, pp. 123–131. Springer, Berlin / Heidelberg, Germany.

ja: Yasuhiro Kawata, Julia Bartels (2000): *Stylebook for the Japanese Treebank in Verbmobil, Report 240*. Universität Tübingen, Tübingen, Germany.

nl: Leonoor van der Beek, Gosse Bouma, Jan Daciuk, Tanja Gaustad, Robert Malouf, Gertjan van Noord, Robbert Prins, Begoña Villada (2002): Chapter 5. The Alpino Dependency Treebank. In *Algorithms for Linguistic Processing NWO PIONIER Progress Report*. Groningen, Netherlands.

pt: Susana Afonso, Eckhard Bick, Renato Haber, Diana Santos (2002): "Floresta sintá(c)tica": A Treebank for Portuguese. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*. ELDA, Las Palmas, Spain.

ro: Mihaela Călăcean (2008): *Data-driven Dependency Parsing for Romanian*. Uppsala Universitet, Uppsala, Sweden.

ru: Igor Boguslavsky, Svetlana Grigorieva, Nikolai Grigoriev, Leonid Kreidlin, Nadezhda Frid (2000): Dependency Treebank for Russian: Concept, Tools, Types of Information. In *Proceedings of the 18th Conference on Computational Linguistics, vol. 2*, pp. 987–991. ACL, Morristown, NJ, USA.

sk: Mária Šimková, Radovan Garabík (2006): Синтаксическая разметка в Словацком национальном корпусе. In *Труды международной конференции Корпусная лингвистика – 2006*, pp. 389–394. ISBN 5-288-04181-4. St. Petersburg University Press, Sankt-Peterburg, Russia.

sl: Sašo Džeroski, Tomaž Erjavec, Nina Ledinek, Petr Pajas, Zdeněk Žabokrtský, Andreja Žele (2006): Towards a Slovene Dependency Treebank. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, pp. 1388–1391. ELDA, Genova, Italy.

sv: Joakim Nivre, Jens Nilsson, Johan Hall (2006): Talbanken05: A Swedish Treebank with Phrase Structure and Dependency Annotation. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*. ELDA, Genova, Italy.

ta: Loganathan Ramasamy, Zdeněk Žabokrtský (2012): Prague Dependency Style Treebank for Tamil. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*. ELDA, İstanbul, Turkey.

tr: Nart B. Atalay, Kemal Oflazer, Bilge Say (2003): The Annotation Process in the Turkish Treebank. In *Proceedings of the 4th International Workshop on Linguistically Interpreted Corpora (LINC)*. EACL, Budapest, Hungary.